Amendments to the Specification:

Please insert the substitute Sequence Listing being filed concurrently herewith into the specification.

On page 4, starting at line 30, please replace the descriptions of Fig.3 and Fig.4 with the following:

Fig. 3 is a map of the cai gene for the CAI protein and summary of the clones used to identify and sequence this gene. In the middle of Fig. 3, upstream of the D3 box, two short peptide sequences are shown: "NEPIYA" (SEQ ID NO:29) and "EEPIYA" (SEQ ID NO:30). At the bottom of the Fig. 3, the nucleotide (SEQ ID NO:11) and deduced amino acid sequence (SEQ ID NO:12) of the cloned segment is shown with peptides D1 (SEQ ID NO:14), D2 (SEQ ID NO:16) and D3 (SEQ ID NO:17) shown boxed.

Figs. 4A through 4F (SEQ ID NO:4 and SEQ ID NO:5) the nucleotide and amino acid sequences of the CAI antigen. The numbers along the left hand margins of Figs. 4A, 4C and 4E designate the amino acid positions. Shown boxed in Fig. 4C-D are two repeats of the peptide EFKNGKNKDFSK (SEQ ID NO:9), which are encoded by the nucleic acid sequence of SEQ ID NO:19. Also shown boxed in Fig. 4C-D are two repeats of the peptide EPIYA (SEQ ID NO:10), the first of which is encoded by the nucleic acid sequence of SEQ ID NO:20, the second of which is encoded by the nucleic acid sequence of SEQ ID NO:21. Also shown boxed in Fig. 4C-D is the peptide FPLKRHDKVDDLSKV (SEQ ID NO:28), which is encoded by the nucleic acid sequence of SEQ ID NO:22.

The paragraph beginning at line 12 on page 6 is replaced with the following:

The "Cytotoxin Associated Immunodominant" (CAI) antigen refers to that protein, and fragments thereof, whose amino acid sequence is described in FIG. 4 and derivatives thereof. The CAI antigen is ~~approximately~~ about 130 kDa as determined by ~~SDS-PAGE~~

SDS/polyacrylamide gel electrophoresis and comprises the following amino acid sequence (SEQ ID NO:25):

```
1   LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGluAsn  20
21  SerValLysAspValIleIleAsnGlnLysValThrAspLysValAspAsnLeuAsnGln  40
41  AlaValSerValAlaLysAlaThrGlyAspPheSerArgValGluGlnAlaLeuAlaAsp  60
61  LeuLysAsnPheSerLysGluGlnLeuAlaGlnGlnAlaGlnLysAsnGluSerLeuAsn  80
81  AlaArgLysLysSerGluIleTyrGlnSerValLysAsnGlyValAsnGlyThrLeuVal 100
101 GlyAsnGlyLeuSerGlnAlaGluAlaThrThrLeuSerLysAsnPheSerAspIleLys 120
121 LysGluLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnGlyLeuLysAsn 140
141 GluProIleTyrAlaLysValAsnLysLysLysAlaGlyGlnAlaAlaSerLeuGluGlu 160
161 ProIleTyrAlaGlnValAlaLysLysValAsnAlaLysIleAspArgLeuAsnGlnIle 180
181 AlaSerGlyLeuGlyValValGlyGlnAlaAlaGlyPheProLeuLysArgHisAspLys 200
201 ValAspAspLeuSerLysValGlyLeuSerArgAsnGlnGluLeuAlaGlnLysIleAsp 220
221 AsnLeuAsnGlnAlaValSerGlu  228
```

SEQ ID NO:25 is the ~~expression product of the following cloned nucleotide sequence (SEQ ID NO:26, uppercase letters only) which entire fragment is cloned into an EcoRI site (EcoRI site in lowercase letters; the entire fragment is shown below as SEQ ID NO:27~~protein encoded by the nucleotides 7 to 691 of the sequenced DNA having the following nucleotide sequence of SEQ ID NO:27, wherein the uppercase letters represent the cloned nucleotide sequence of SEQ ID NO:26 and the lowercase letters represent the EcoRI site:

```
1   gaattcAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAAGCAAAAAGCGACCTT 60
61  GAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGATAAAGTTGATAATCTC 120
121 AATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTCAGTAGGGTAGAGCAAGCGTTA 180
181 GCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCCAACAAGCTCAAAAAAATGAAAGT 240
241 CTCAATGCTAGAAAAAAATCTGAAATATATCAATCCGTTAAGAATGGTGTGAATGGAACC 300
301 CTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCTAAAAACTTTTCGGAC 360
361 ATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAATAACAATAATGGACTC 420
421 AAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGGCAAGCAGCTAGCCTT 480
481 GAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAAT 540
541 CAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCAT 600
601 GATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAATTGGCTCAGAAA 660
661 ATTGACAATCTCAATCAAGCGGTATCAGAAGccgaattc 699
```

The paragraph beginning at line 15 on page 52 is replaced with the following:

The cai gene coded for a putative protein of 1147 amino acids, with predicted molecular weight of 128012.73 Daltons and an isoelectric point of 9.72. The basic properties

of the purified protein were confirmed by two dimensional gel electrophoresis. The codon usage and the GC content (37%) of the gene were similar to that described for other *H. pylori* genes (13,26). A putative ribosome binding site: AGGAG, was identified 5 base pairs upstream from the proposed ATG starting codon. Computer search for promoter sequences of the region upstream from the ATG start codon, identified sequences resembling either -10 or -35 regions, however, a region with good consensus to an *E. coli* promoter, or resembling published *H. pylori* promoter sequences was not found. Primer extension analysis of purified *H. pylori* RNA showed that 104 and 214 base pairs upstream from the ATG start codon there are two transcriptional start sites. Canonical promoters could not be identified upstream from either transcriptional initiation sites. The expression of a portion of the CAI antigen by clone 57/D suggests that *E. coli* is also recognizing a promoter in this region, however, it is not clear whether *E. coli* recognizes the same promoters of *H. pylori* or whether the *H. pylori* DNA that is rich in A-T provides E. coli with regions that may act as promoters. A rho independent terminator was identified downstream from the stop codon. In FIG. 4, the AGGAG ribosome binding site and terminator are underlined, and the repeated sequence and motif containing 6 asparagines (SEQ ID NO:23) are boxed. The CAI antigen was very hydrophilic, and did not show obvious leader peptide or transmembrane sequences. The most hydrophilic region was from amino acids 600 to 900, where also a number of unusual features can be observed: the repetition of the sequences EFKNGKNKDFSK (SEQ ID NO:9) and EPIYA (SEQ ID NO:10), and the presence of a stretch of six contiguous asparagines (boxed in FIG. 4)(SEQ ID NO:23) which is encoded by the sequence of SEQ ID NO:24.

The paragraph beginning at line 11 on page 53 is replaced with the following:

Diversity of the gene appears to be generated by internal duplications. To find out the mechanism of size heterogeneity of the CAI proteins in different strains, the structure of one of the strains with a larger CAI protein (G39) was analyzed using Southern blotting, PCR and DNA sequencing. The results showed that the cai gene of G39 and CCUG 17874 were identical in size until position 3406, where the G39 strain was found to contain an insertion of 204 base pairs, made by two identical repeats of 102 base pairs. Each repeat was found to

contain sequences deriving from the duplication of 3 segments of DNA (sequences D1 (SEQ ID NO:13), D2 (SEQ ID NO:15) and D3 (SEQ ID NO:18) in FIG. 3) coming from the same region of the cai gene and connected by small linker sequences. A schematic representation of the region where the insertion occurred and of the insertion itself is shown in FIG. 3. The nucleotide sequence of the insertion shown (SEQ ID NO:11) has the deduced amino acid sequence shown (SEQ ID NO:12).